Does microhabitat influence cryptic diversity?: An investigation using microsnails (*Punctum randolphi*) from the Pacific Northwest rainforests.

David J. Sneddon, Megan L. Smith, Bryan C. Carstens

Department of Evolution, Ecology, and Evolutionary Biology. The Ohio State University.

Columbus, Ohio.

Abstract

The Pacific Northwest temperate rainforests are biodiversity hotspots with a dramatic geological history, with many species endemic to forests in the Cascades mountains. The Pleistocene glacial cycles drastically impacted the ranges of these endemics, with the much of the Northern Cascades mountain range being covered by alpine ice. Previous work suggests that many species survived the Pleistocene in small isolated refugia throughout the Pleistocene glaciation. Here, we investigate whether the microsnail *Punctum randolphi* was able to survive in the Northern Cascades throughout the Pleistocene by living in refugia and expand into its current range after the last glacial maximum. We also investigate whether *P. randolphi* harbors cryptic diversity due to divergence in these isolated refugia. Results from gene tree inference, multimodel inference and niche modeling suggest that *P. randolphi* was able to survive throughout Pleistocene glaciation in a single isolated refugium before expanding its range after the last glacial maximum.

Introduction

The mesic forest ecosystems of the Pacific Northwest are a center of biodiversity and speciation (Brunsfeld *et al.,* 2001). The Pacific Northwest harbors two large mesic rainforest ecosystems, one in the coastal Cascades mountain range, and one in the inland Northern Rocky Mountains. These ecosystems are disjunct and separated by more than two hundred kilometers by the Columbia Basin. The Columbia Basin can be characterized as an arid shrub-steppe ecosystem. Prior to the formation of the Columbia Basin, the Pacific Northwest was a continuous mesic forest, with a disjunction forming concurrently to the orogeny of the Cascades Range and subsequent rain shadow (Graham, 1999). The Cascades restricted rainfall between the two ranges forming the arid Columbia Basin. The shrub-steppe ecosystem of the Columbia Basin is not suitable for the many rainforest endemic species that reside in the rainforests of the Cascades and Northern Rocky Mountains and has proven to act as a barrier to dispersal for rainforest endemic organisms (Carstens *et al.,* 2005).

Pleistocene glaciation disrupted the ranges and distributions of many Pacific Northwest species (Pielou, 2008). During this glacial period, the Pacific Northwest rainforests were repeatedly covered by advancing sheets of alpine ice for periods of 90,000 years at a time. During these periods, many species were forced south out of their ranges to the unglaciated part of the Pacific Northwest. However, many of the rainforest endemics in the coastal Cascades Mountain range were able to persist in isolated refugia, often located in deep river canyons (Barnosky *et al.,* 1987).

The presence of these isolated refugia leads to three different distribution patterns of Pacific Northwest species (Brunsfeld *et al.,* 2001). The first pattern implies that there was no change in the distribution of organisms throughout the Pleitocene glacial cycles. Under this

pattern, we posit that species ranges persisted throughout glacial maxima and does not predict isolated refugia within the glacial area. The second pattern implies that a single isolated refugium was present throughout Pleistocene glaciation. Under this assumption, rainforest endemics were able to survive in a single small isolated area and expanded their ranges after the retreat of the glaciers. Under this model, we would not predict any significant divergence between populations. The third pattern of distribution implies that instead of only a single isolated refugium, there were multiple refugia (Soltis *et al.,* 1997). Support for this distributional pattern would predict genetic divergence between populations from separate refugia leading to the presence of undocumented cryptic diversity, or deep genetic divergence between organisms that are identical with regard to their morphology (Bickford *et al.,* 2007).

One area of the rainforest that harbors a large amount of invertebrate diversity is the forest floor, specifically within the wet leaf litter that collects on the floor. This leaf litter is home to the microsnail *Punctum randolphi*. *P. randolphi* is a small pulmonate snail with a shell diameter measuring 1-1.4 mm that is endemic to the rainforests of the Cascades and Northern Rocky Mountains. Here, we use molecular and environmental data to investigate whether *P. randolphi* was able to survive the Pleistocene glacial cycles within multiple isolated refugia or a single refugium, and we investigate whether this species harbors undescribed cryptic diversity.

Materials and Methods

Sampling method

P. randolphi are efficiently collected by gathering wet leaf litter from the forest floor. We collected 292 samples of leaf litter from 146 different sites within the rainforests of the Cascades. Wet leaf litter samples were frozen immediately to prevent degradation of organisms.

Microinvertebrates were removed from leaf litter after all leaf litter samples were collected. Each leaf litter sample was sorted through for 45 minutes under light and 10x magnification. Whenever an invertebrate was found, the organism was removed from the leaf litter and immediately placed into 95% ethanol. Since destructive sampling was necessary to obtain DNA due to the small size of microsnails, photos were taken of each microsnail prior to DNA extraction. We removed 16 *Punctum randolphi* individuals from the leaf litter from 10 different sites.

DNA isolation and sequencing

DNA was extracted using DNeasy Blood and Tissue kits (Quiagen). *Punctum randolphi* and other microsnails are too small to preserve any of the organisms' tissue during extraction. Microsnails are also not operculate, meaning they don't have a hard plate that seals off the opening of their shell, therefore we placed the entire organism into ATL buffer and Proteinase K solution and placed them into a 56ºC water bath. Samples were in the water bath for 90 minutes. Once removed, any remaining shell from any samples were removed prior to advancing. The remainder of the DNA extraction was performed using the manufacturer's standard protocol. A 710-bp portion of the COI gene was amplified via PCR using the HCO2198 and LCO1490 forward and reverse primers (Folmer *et al.,* 1994) and sequenced using the Sanger sequencing method (Sanger *et al.,* 1977). Sequences were edited by hand when necessary using Geneious Prime v2019.0.4 (Kearse *et al.,* 2012). Sequence alignment was preformed using MUSCLE alignment algorithm (Edgar, 2004) implemented within Geneious using default settings. Microsnails were identified by searching the NCBI nucleotide database using the blastn program implemented within Geneious (McGinnis & Madden, 2004). In addition to the data collected for

this study, 15 available *Punctum* COI sequences were retrieved from GenBank in March 2019. Additionally, a sequence from one species within Punctidae, but not from the *Punctum* genus (*Paralaoma servilis*), and one species from outside of Punctidae (*Discus catskillensis*) were downloaded from GenBank and used as outgroups.

Gene tree estimation

We used Neighbor Joining search using BioNJ (Gascuel, 1997) method calculated from Jukes-Cantor distance settings to obtain a starting tree. Automated Model Selection function (implemented within PAUP*) was used to select the best model of analysis, with 11 substitution schemes, gamma rate variation, and a proportion of invariant sites considered. All models were evaluated using AICc, Decision Theory, and BIC. Once a model of sequence evolution was identified, it was used to estimate a gene tree using GARLI v0.951 (Zwickl, 2006) with 100 bootstrap replicates. MrBayes v3.2.6 (Ronquist *et al.,* 2003) was used to infer a posterior distribution of gene trees under this same model. We ran two independent runs with 4,000,000 generations and four chains per run and four rate categories. Tracer v1.7 (Rambaut *et al.,* 2018) was used to visualize results and assess convergence. We discarded 25% of the samples as burn-in and summarized our results using a majority rule consensus tree

Approximate Bayesian Computation

Given the results from Bayesian and Maximum Likelihood gene tree estimates, we tested two models of historical demography using Approximate Bayesian Computation. Only two models were tested because the lack of divergence within *P. randolphi* does not support a multiple refugia model. We simulated data under the constant population model that suggests that Pleistocene glacial cycles did not impact distribution and population of *P. randolphi*, and

under an expansion model, which suggests that *P. randolphi* was able to survive throughout the most recent glacial maximum within a single refugium, then expand into its current distribution. Data were simulated under each model using Hudson's ms (Hudson, 2002). Theta was drawn from a uniform (0.001, 15) prior in units of $4*N_0*\mu$. The timing of expansion was drawn from a uniform (0.375, 0.750) prior in units of $4*N_0$*generation time. Growth rate was drawn from a uniform (5, 20) prior. We simulated 100,000 datasets in total. We calculated the summary statistics pi, the number of segregating sites, Tajima's D, and Fay and Wu's H using the program sample stats (Hudson, 2002). We calculated the same summary statistics from our empirical data using the R package PopGenome and calculated the posterior probabilities of each model. We used a simple rejection step and threshold of 0.001.

Species Distribution Modeling

We used sampling localities from this study and GenBank, as well as data downloaded from GBIF (GBIF, 2019) to construct a species distribution model (SDM). We created a species distribution model in R using maximum entropy to assess areas that are suitable for *P. randolphi* within the Pacific Northwest both currently and during the last glacial maximum (~20,000 year ago). The model was created using the maxent command (Elith *et al.,* 2011) implemented in the R package "biomod2" (Thuiller *et al.,* 2016) using 10,000 background points. Bioclim variables were obtained from Worldclim (Hijmans *et al.,* 2005) database under current conditions. We used ccsm4 data from worldclim for climatic variables during last glacial maximum. Using our species distribution model and our climate variables from worldclim, we projected the SDM to a map of both present and Last Glacial Maximum conditions to predict habitat suitability for *P. randolphi*.

Results

Gene tree estimation

COI sequence data was collected from 16 *P. randolphi* individuals collected from leaf litter. An additional 15 *Punctum* COI sequences and 2 outgroup sequences were downloaded from GenBank. Our final alignment was 655 base pairs long. AutoModel results from PAUP* suggested the TPM3uf+G model under corrected AIC, BIC and Decision Theory. Maximum likelihood and Bayesian gene trees were created using the closest models available in GARLI and mrbayes, which included 6 rate parameters instead of 3 (Figure 2). Most *P. randolphi* belonged to a single monophyletic group, though 3 GenBank samples identified as *P. randolphi* formed a strongly supported clade. There was little support for the relationship of this clade to other sampled *P. randolphi*. There were no other deep divergences within *P. randolphi*. These results were supported by the Maximum Likelihood inference (Figure 3).

Approximate Bayesian Computation

We did not observe deep genetic divergence between *P. randolphi* individuals therefore we do not have any support for the multiple refugia hypothesis. Our empirical data estimated nucleotide diversity (π) of 6.32971, and contained 30 segregating sites. Tajima's D was estimated to be -0.7997 and Fay and Wu's H to be 0.04641. The range of variation in our prior distribution contained these values. Model 1 (no change in population structure) had a posterior probability of 0.27, while Model 2 (expansion) had a posterior probability of 0.73.

Species Distribution Modeling

Current distribution models suggested suitable habitat across entire coastal Cascades and inland Northern Rocky mountain ranges (Figure 4). Sampling localities both from our own sampling and from data available on GBIF all fell within present day habitat indicated by our model (AUC=0.934). Our species distribution model suggests very little suitable habitat for *P. randolphi* during the last glacial maximum (Figure 5). Small pockets of suitable habitat were found both in the coastal and inland ranges with a single hotspot in the coastal range.

Discussion

Our species distribution model supports the theory that *P. randolphi* was able to survive in a single isolated refugium throughout the last glacial maximum. The single area of the coastal range indicated on our map with high suitability supports the theory that there were not multiple refugia for *P. randolphi.* This is also supported by the lack of genetic divergence within *P. randolphi* as indicated in our tree. Alternative explanations for this tree could be due to the possibility that, despite the fact that the COI gene is a mitochondrial gene, it might not have as high of a substitution rate as we thought. Repeating this study with a different mitochondrial gene could provide more support for our interpretation. Results from our Approximate Bayesian Computation support the expansion model, suggesting that *P. randolphi* expanded from this refugium to its current range following the last glacial maximum.

The lack of genetic diversity indicated by our tree combined with such a large range indicates that there is migration occurring between *P. randophi* populations. We did not find that *P. randolphi* is geographically distinct. The way that these microsnails are facilitating their movement remains unknown, although there has been evidence for snails dispersing across geographic barriers through the digestive tract of birds (Wilke, Duncan, 2004).

Similar results have been found in other related studies. Other terrestrial Gastropod species have shown very little cryptic diversity (Smith *et al.,* 2018). However, these results did predict more population structure within the Cascades mountain range than we see with *P. randolphi*. This lack of cryptic diversity can also be seen in more distant species. Certain Microtus species have also displayed this lack of cryptic diversity in the Pacific Northwest rainforests. The presence of isolated refugia within the Cascades mountain range have also been observed in multiple plant species (Soltis *et al.,* 1997) as well as amphibian species (Steel *et al.,* 2006). These isolated refugia enabled survival throughout Pleistocene glacial cycles, and allowed for dispersal after the last glacial maximum.

References

Barnosky, C. W., Anderson, P. M., & Bartlein, P. J. (1987). The northwestern US during deglaciation; vegetational history and paleoclimatic implications. North America and adjacent oceans during the last deglaciation, 3, 289-321.

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2008). GenBank. Nucleic Acids Research, 37(suppl_1), D26-D31.

Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K., Meier, R., Winker, K., ... & Das, I. (2007). Cryptic species as a window on diversity and conservation. Trends in Ecology & Evolution, 22(3), 148-155.

Brunsfeld, S. J., Sullivan, J., Soltis, D. E., & Soltis, P. S. (2001). Comparative phylogeography of northwestern North America: a synthesis. Special Publication-British Ecological Society, 14, 319-340.

Carstens, B. C., Brunsfeld, S. J., Demboski, J. R., Good, J. M., & Sullivan, J. (2005). Investigating the evolutionary history of the Pacific Northwest mesic forest ecosystem: hypothesis testing within a comparative phylogeographic framework. Evolution, 59(8), 1639-1652.

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research, 32(5), 1792-1797.

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. Diversity and Distributions, 17(1), 43-57.

Espíndola, A., Ruffley, M., Smith, M. L., Carstens, B. C., Tank, D. C., & Sullivan, J. (2016, October). Identifying cryptic diversity with predictive phylogeography. In Proc. R. Soc. B (Vol. 283, No. 1841, p. 20161529). The Royal Society.

Folmer O, Black M, Hoeh W, Lutz R & Vrijenhoek R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology 3, 294–299

Gascuel, O. (1997). BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. Molecular Biology and Evolution, 14(7), 685-695.

GBIF.org (2019), *GBIF Home Page*. Available from: https://www.gbif.org [13 August 2018].

Graham, A. 1999. Late Cretaceous and Cenozoic history of North American vegetation. Oxford Univ. Press, New York.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology, 25(15), 1965-1978.

Hudson, R. R. (2002). Generating samples under a Wright–Fisher neutral model of genetic variation. Bioinformatics, 18(2), 337-338.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., ... & Thierer, T. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics, 28(12), 1647-1649.

McGinnis, S., & Madden, T. L. (2004). BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic acids research, 32(suppl_2), W20-W25.
Pfeifer, Bastian, et al. "PopGenome: an efficient Swiss army knife for population genomic analyses in R." Molecular Biology and Evolution 31.7 (2014): 1929-1936.

Pielou E. C. (2008). After the ice age: the return of life to glaciated North America. Chicago: University of Chicago Press

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Systematic Biology, 67(5), 901-904.

Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics, 19(12), 1572-1574.

Sanger, Frederick, Steven Nicklen, and Alan R. Coulson. “DNA sequencing with chain-terminating inhibitors.” Proceedings of the National Academy of Sciences 74.12 (1977): 5463-5467.

Smith, M. L., Ruffley, M., Rankin, A. M., Espíndola, A., Tank, D. C., Sullivan, J., & Carstens, B. C. (2018). Testing for the presence of cryptic diversity in tail-dropper slugs (Prophysaon) using molecular data. Biological Journal of the Linnean Society, 124(3), 518-532.

Soltis, D. E., Gitzendanner, M. A., Strenge, D. D., & Soltis, P. S. (1997). Chloroplast DNA intraspecific phylogeography of plants from the Pacific Northwest of North America. Plant Systematics and Evolution, 206(1-4), 353-373.

Steele, C. A., & Storfer, A. (2006). Coalescent-based hypothesis testing supports multiple Pleistocene refugia in the Pacific Northwest for the Pacific giant salamander (Dicamptodon tenebrosus). Molecular Ecology, 15(9), 2477-2487.

Swofford, D. L. (2001). Paup*: Phylogenetic analysis using parsimony (and other methods) 4.0. B5.

Thuiller, Wilfried, et al. ”Package ‘biomod2’.” Species distribution modeling within an ensemble forecasting framework (2016).

Wilke, T., & Duncan, N. (2004). Phylogeographical patterns in the American Pacific Northwest: lessons from the arionid slug Prophysaon coeruleum. Molecular Ecology, 13(8), 2303-2315.

Zwickl, D. J. (2006). GARLI: genetic algorithm for rapid likelihood inference. See http://www. Bio. Utexas. Edu/faculty/antisense/garli/Garli. Html.

Figures

Figure 1: Species distribution of *P. randolphi* shown on a map of the Pacific Northwest. The inset depicts the shell morphology of this species with a 1mm scale bar.

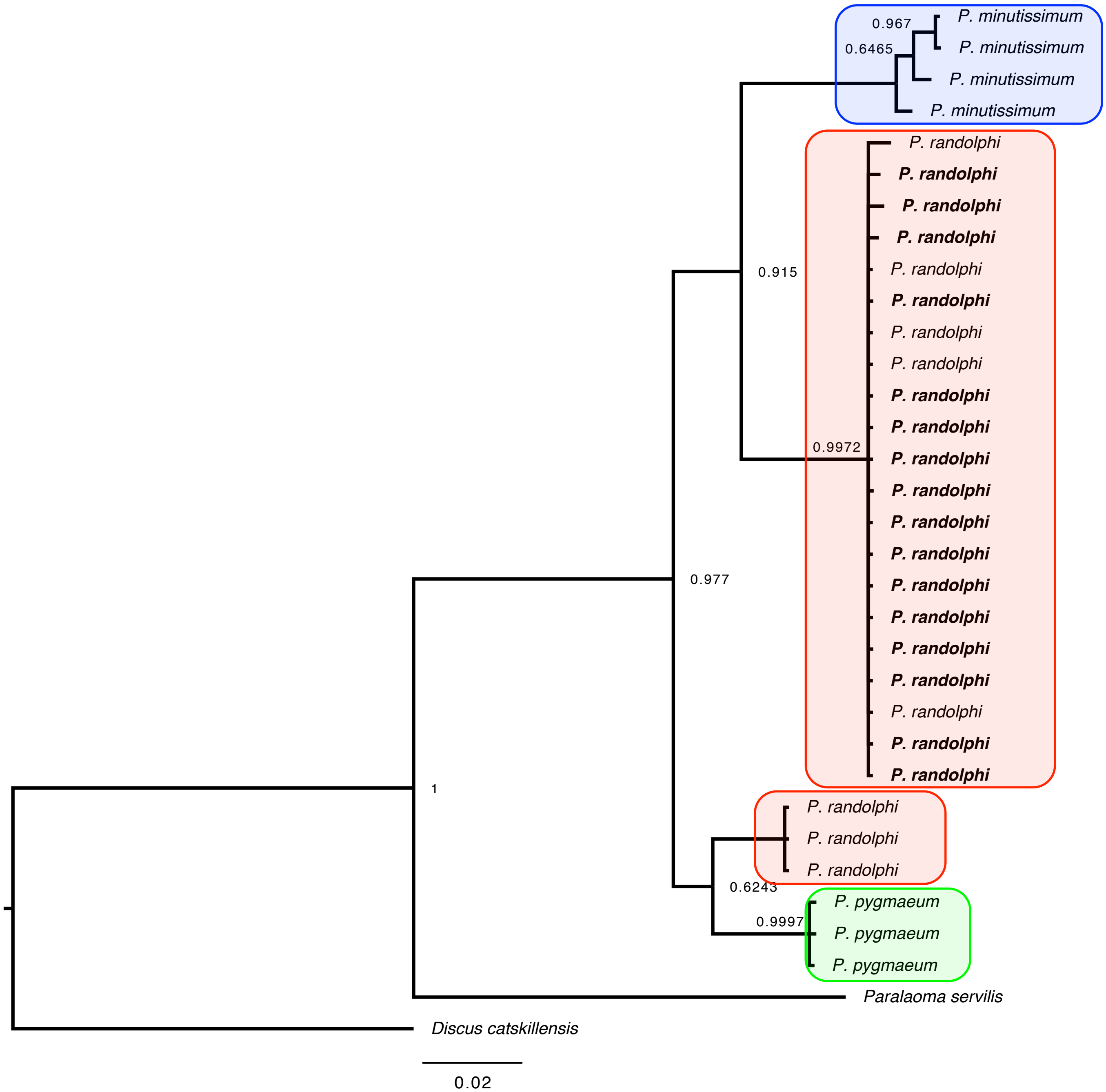

Figure 2: Bayesian COI gene tree of *Punctum* samples created using MrBayes. Node labels indicate probability. Bold text indicates samples collected for this study.

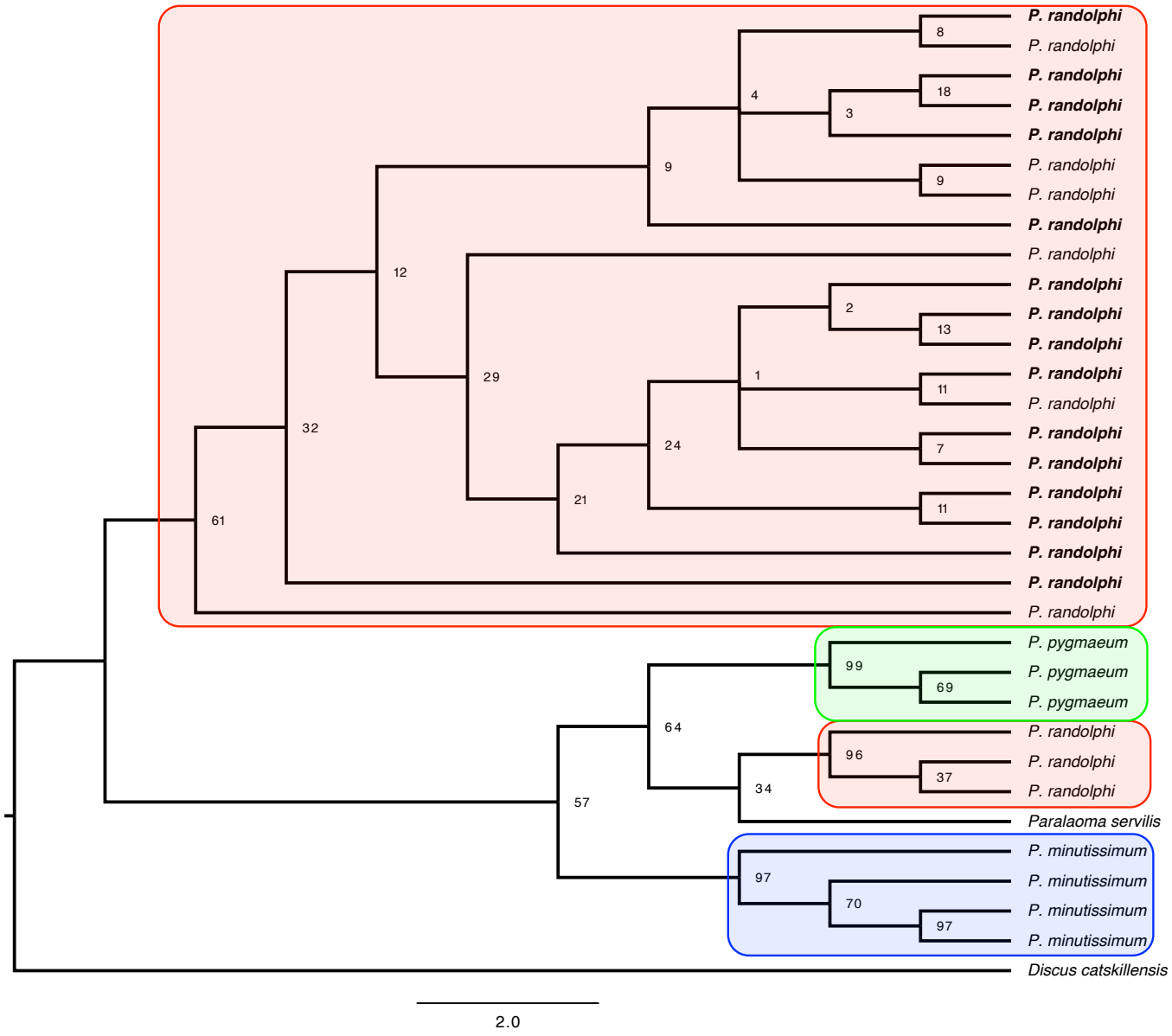

Figure 3: Maximum Likelihood COI gene tree of *Punctum* individuals created using GARLI. Node labels indicate bootstrap score. Bold text indicates samples collected for this study.

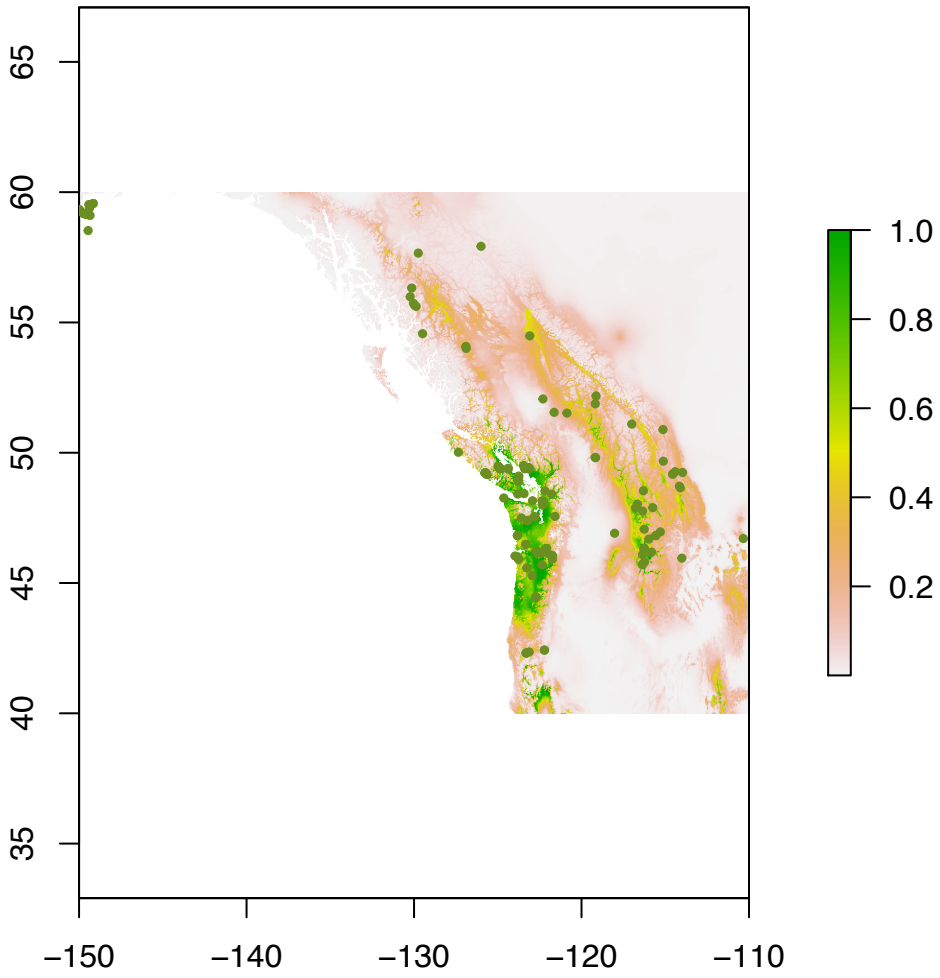

Figure 4: Current suitable habitat for *P. randolphi* as indicated by our Species Distribution Model. Scale represents increasing suitability.

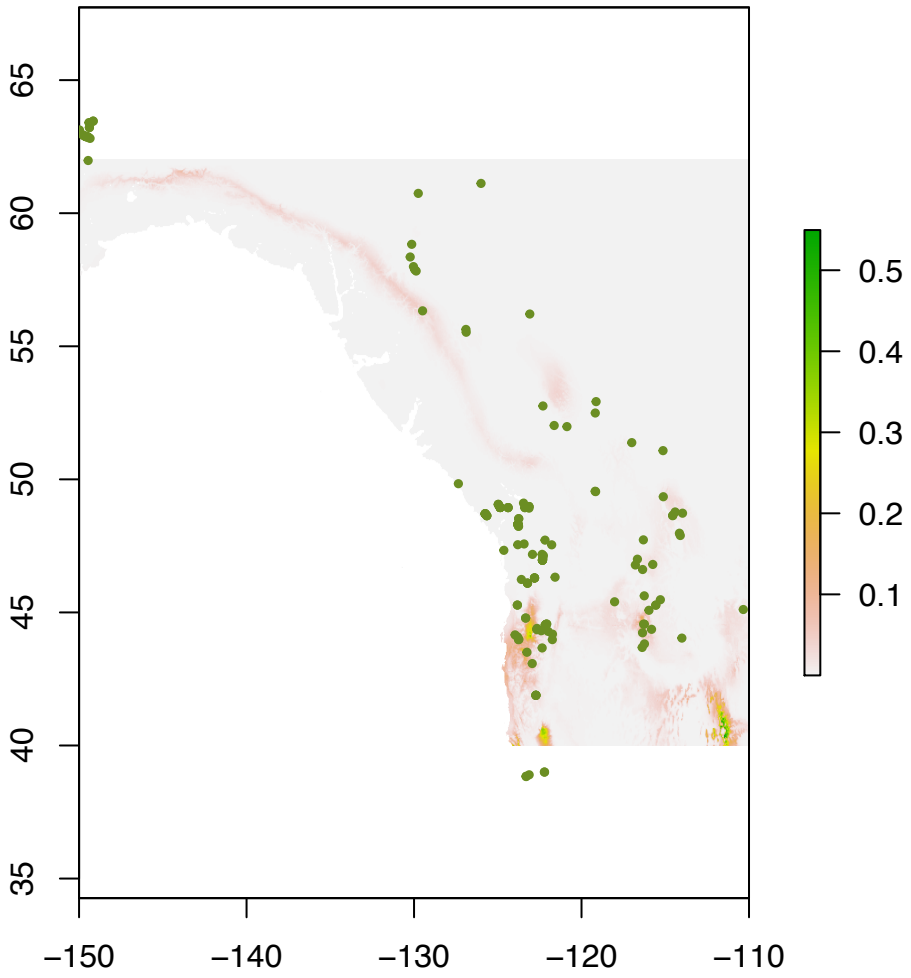

Figure 5: Suitable habitat during the last glacial maximum for *P. randolphi* as indicated by our Species Distribution Model. Scale represents increasing suitability.